

# Statistical inference for max-stable processes in space and time

Richard A. Davis <sup>\*</sup>      Claudia Klüppelberg <sup>†</sup>      Christina Steinkohl <sup>‡</sup>

March 3, 2013

## Abstract

Max-stable processes have proved to be useful for the statistical modelling of spatial extremes. Several representations of max-stable random fields have been proposed in the literature. One such representation is based on a limit of normalized and scaled pointwise maxima of stationary Gaussian processes that was first introduced by Kabluchko, Schlather and de Haan [14].

This paper deals with statistical inference for max-stable space-time processes that are defined in an analogous fashion. We describe pairwise likelihood estimation, where the pairwise density of the process is used to estimate the model parameters and prove strong consistency and asymptotic normality of the parameter estimates for an increasing space-time dimension, i.e., as the joint number of spatial locations and time points tends to infinity. A simulation study shows that the proposed method works well for these models.

*AMS 2010 Subject Classifications:* primary: 60G70  
secondary: 62F12, 62M10, 62M40

*Keywords:* Max-stable space-time process, pairwise likelihood estimation, strong consistency, asymptotic normality

---

<sup>\*</sup>Department of Statistics, Columbia University, New York, United States,  
<http://www.stat.columbia.edu/>, Email: [rdavis@stat.columbia.edu](mailto:rdavis@stat.columbia.edu)

<sup>†</sup>Center for Mathematical Sciences, Technische Universität München, D-85748 Garching, Germany,  
<http://www-m4.ma.tum.de>, Email: [cklu@ma.tum.de](mailto:cklu@ma.tum.de)

<sup>‡</sup>Center for Mathematical Sciences and Institute for Advanced Study, Technische Universität München, D-85748 Garching, Germany, <http://www-m4.ma.tum.de>, Email: [steinkohl@ma.tum.de](mailto:steinkohl@ma.tum.de)

# 1 Introduction

Max-stable processes have proven to be useful in the modelling of spatial extremes. Typically, meteorological extremes like heavy rainfall or extreme wind speeds are modelled using extreme value theory. In particular, sample maxima such as annual maximum wind speeds are observed at several locations of some spatial process. Other applications may involve the analysis of image data resulting from tomographic examinations.

Several representations of max-stable processes have been proposed in the literature, including for example Brown and Resnick [3], de Haan [8], Kabluchko, Schlather and de Haan [14], and Schlather [19]. Recently, models for extreme values observed in a space-time setting have generated a great deal of interest. First approaches can be found in Davis and Mikosch [6], Huser and Davison [11], Kabluchko [13], and Davis, Klüppelberg and Steinkohl [5].

In this paper, we follow the approach described in Davis et al. [5], who extend the max-stable process introduced in Kabluchko et al. [14] to a space-time setting. The process is constructed as the limit of rescaled and normalized maxima of independent replications of some stationary Gaussian space-time process. The underlying correlation function of the Gaussian process is assumed to belong to a parametric model whose parameters describe smoothness of the correlation function near the origin.

As it is well-known for max-stable processes, the full likelihood function is computationally intractable and other methods have to be used to derive parameter estimates. Standard procedures for such cases are composite likelihood including pairwise likelihood estimation. These methods go back to Besag [1], and there is an extensive literature available dealing with applications and properties of the estimates, see for example Cox and Reid [4], Lindsay [16], Varin [23], or Varin and Vidoni [24]. Recent work concerning the application of pairwise likelihood methods to max-stable random fields can be found in Huser and Davison [11] and Padoan, Ribatet and Sisson [18].

Since the observations in a space-time setting are correlated in space and time, we use special properties of max-stable processes to show strong consistency and asymptotic normality of the estimates. Here, it is assumed that the locations lie on a regular lattice and that the time points are equidistant. The spatial and/or the temporal dimension, i.e., the number of spatial locations and/or time points, increases to infinity. The main step of the proof is based on a strong law of large numbers for the pairwise likelihood function. Stoev [20] analyzed ergodic properties for max-stable processes in time resulting from extremal integral representations for max-stable processes that were introduced in Stoev and Taquq [21]. The extension to a spatial setting and the resulting strong law of large numbers was shown by Wang, Roy and Stoev [26]. By combining these two results we obtain a strong law of large numbers for a jointly increasing space-time domain.

In addition to strong consistency, we prove asymptotic normality for the pairwise likelihood estimates. A first result concerning asymptotic normality of pairwise likelihood estimates for max-stable space-time processes can be found Huser and Davison [11], who fix the number of locations and let the number of time points tend to infinity. We formulate asymptotic normality for the space-time setting and use Bolthausen's theorem [2] together with strong mixing properties shown by Dombry and Eyi-Minko [9] to prove asymptotic normality for an increasing number of space-time locations.

Our paper is organized as follows. In Section 2, we introduce the max-stable space-time process for which inference properties will be considered in subsequent sections. Section 3.2 describes pairwise likelihood estimation and the particular setting for our model. In Section 4 we establish strong consistency for the estimates for increasing space-time domain. Asymptotic

normality of these parameters is established in Section 5. A simulation study evaluating the performance of the estimates is described in Section 6.

## 2 Description of the model

We start with the process that will be used for modelling extremes in space and time; details can be found in Davis et al. [5]. Let  $\{Z(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in [0, \infty)\}$  denote a stationary space-time Gaussian process on  $\mathbb{R}^d \times [0, \infty)$  with mean zero and variance one. With the correlation function

$$\rho(\mathbf{h}, u) = \mathbb{E}[Z(\mathbf{s}, t)Z(\mathbf{s} + \mathbf{h}, t + u)],$$

where  $\mathbf{h} \in \mathbb{R}^d$  is the spatial lag and  $u \in \mathbb{R}$  is the time lag, we make the following assumption that will be used throughout the paper.

**Assumption 2.1.** *There exist sequences of constants  $s_n \rightarrow 0$ ,  $t_n \rightarrow 0$  as  $n \rightarrow \infty$ , such that*

$$\log n(1 - \rho(s_n \mathbf{h}, t_n u)) \rightarrow \delta(\mathbf{h}, u) > 0, \quad \text{as } n \rightarrow \infty.$$

Assumption 2.1 is natural in the context of stationary space-time models; the correlation function tends to one at a certain rate as the space-time lag approaches the zero.

**Proposition 2.2** (Kablichko et al. [14] and Davis et al. [5]). *Let  $\{Z_j(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in [0, \infty)\}$ ,  $j = 1, \dots, n$ , be independent replications of the space-time Gaussian process described above and let  $\{\xi_j, j \in \mathbb{N}\}$  denote points of a Poisson random measure on  $[0, \infty)$  with intensity measure  $\xi^{-2}d\xi$ . Suppose Assumption 2.1 is satisfied. Then, the random fields  $\{\eta_n(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in [0, \infty)\}$ , defined for  $n \in \mathbb{N}$  by*

$$\eta_n(\mathbf{s}, t) = \bigvee_{j=1}^n -\frac{1}{\log(\Phi(Z_j(s_n \mathbf{s}, t_n t)))}, \quad \mathbf{s} \in \mathbb{R}^d, t \in [0, \infty), \quad (1)$$

*converge weakly on the space of continuous functions on  $\mathbb{R}^d \times [0, \infty)$  to the stationary Brown-Resnick process*

$$\eta(\mathbf{s}, t) = \bigvee_{j=1}^{\infty} \xi_j \exp\{W_j(\mathbf{s}, t) - \delta(\mathbf{s}, t)\}, \quad (2)$$

*where the deterministic function  $\delta$  is given in Assumption 2.1 and  $\{W_j(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in [0, \infty)\}$ ,  $j \in \mathbb{N}$  are independent replications of a Gaussian process with stationary increments,  $W(\mathbf{0}, 0) = 0$ ,  $\mathbb{E}(W(\mathbf{s}, t)) = 0$  and covariance function for  $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^d, t_1, t_2 \in [0, \infty)$*

$$\text{Cov}(W(\mathbf{s}_1, t_1), W(\mathbf{s}_2, t_2)) = \delta(\mathbf{s}_1, t_1) + \delta(\mathbf{s}_2, t_2) - \delta(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2).$$

*The bivariate distribution function of  $\eta$  can be expressed in closed form and is based on a well-known result by Hüsler and Reiss [12];*

$$F(x_1, x_2) = \exp \left\{ -\frac{1}{x_1} \Phi \left( \frac{\log \frac{x_2}{x_1}}{2\sqrt{\delta(\mathbf{h}, u)}} + \sqrt{\delta(\mathbf{h}, u)} \right) - \frac{1}{x_2} \Phi \left( \frac{\log \frac{x_1}{x_2}}{2\sqrt{\delta(\mathbf{h}, u)}} + \sqrt{\delta(\mathbf{h}, u)} \right) \right\}, \quad (3)$$

*where  $\Phi$  denotes the distribution function of a standard normal distribution.*

Many correlation functions satisfy the following condition, which will be used throughout.

**Condition 2.3.** *The correlation function has an expansion around zero, given by*

$$\rho(\mathbf{h}, u) = 1 - \theta_1 \|\mathbf{h}\|^{\alpha_1} - \theta_2 |u|^{\alpha_2} + O(\|\mathbf{h}\|^{\alpha_1} |u|^{\alpha_2}), \mathbf{h} \in \mathbb{R}^d, u \in \mathbb{R},$$

where  $0 < \alpha_1, \alpha_2 \leq 2$  and  $\theta_1, \theta_2 > 0$ .

Condition 2.3 allows for an explicit expression of the limit function  $\delta$  in Assumption 2.1,

$$\delta(\mathbf{h}, u) = \theta_1 \|\mathbf{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2}, \quad (4)$$

where the scaling sequences  $(s_n)$  and  $(t_n)$  can be chosen as  $s_n = (\log n)^{1/\alpha_1}$  and  $t_n = (\log n)^{1/\alpha_2}$ . The parameters  $\alpha_1, \alpha_2 \in (0, 2]$  relate to the smoothness of the underlying Gaussian process in space and time, where the case  $\alpha_1 = \alpha_2 = 2$  corresponds to a mean-square differentiable process. For example, Gneiting's class of correlation functions [10] satisfies Condition 2.3. For a detailed analysis of Gneiting's class and further examples we refer to Davis et al. [5], Proposition 4.5, where the expansion around zero is calculated for several classes of correlation functions. A further property of the model defined in Proposition 2.2 is the closed form expression for the *tail dependence coefficient*, which is defined by

$$\chi(\mathbf{h}, u) = \lim_{x \rightarrow \infty} P\left(\eta(\mathbf{s}_1, t_1) > F_{\eta(\mathbf{s}_1, t_1)}^{\leftarrow}(x) \mid \eta(\mathbf{s}_2, t_2) > F_{\eta(\mathbf{s}_2, t_2)}^{\leftarrow}(x)\right),$$

where  $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$  denotes the spatial distance between two locations and  $u = t_1 - t_2$  is the temporal lag. As derived in Section 3 in Davis et al. [5], we obtain

$$\chi(\mathbf{h}, u) = 2 \left(1 - \Phi(\sqrt{\delta(\mathbf{h}, u)})\right) = 2 \left(1 - \Phi(\sqrt{\theta_1 \|\mathbf{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2}})\right). \quad (5)$$

### 3 Pairwise likelihood estimation

In this section, we describe the pairwise likelihood estimation for the parameters of the model in (1) introduced in Section 2. Composite likelihood methods have been used, whenever the full likelihood is not available or intractable. We present the general definition of composite and pairwise likelihood functions in Section 3.1. Afterwards, we describe the details for our model.

#### 3.1 Basics on composite likelihood estimation

Composite likelihood methods go back to Besag [1] and Lindsay [16] and there is vast literature available, from a theoretical and an applied point of view. For more information we refer to Varin [23] who gives an overview of existing models and inference including extensive references. In the most general setting, the composite log-likelihood function is given by

$$l_c(\boldsymbol{\psi}, \mathbf{x}) = \sum_{i=1}^q w_i \log f(\mathbf{x} \in A_i; \boldsymbol{\psi}).$$

From this general form, special composite likelihood functions can be derived. For our setting we define the (*weighted*) *pairwise log-likelihood function* by

$$PL(\boldsymbol{\psi}; \mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n w_{i,j} \log f_{\boldsymbol{\psi}}(x_i, x_j), \quad (6)$$

where  $\mathbf{x} = (x_1, \dots, x_n)$  is the data vector,  $f_\psi(x_i, x_j)$  is the density for the bivariate observations  $(x_i, x_j)$  and  $w_{i,j}$  are weights which can be used for example to reduce the number of pairs included in the estimation. The parameter estimates are obtained by maximizing (6).

As noted in Cox and Reid [4], for dependent observations, estimates based on the composite likelihood need not be consistent or asymptotically normal. This is important for space-time applications, since all components may be highly dependent across space and time.

### 3.2 Application to spatio-temporal max-stable random fields

To derive pairwise-likelihood functions for the model defined in Proposition (1) we first need to derive the bivariate density function for the space-time max-stable process. For later purposes we state the closed form expression in the following lemma. Throughout we denote by  $\Phi$  and  $\varphi$  the cumulative distribution function and the density of the standard normal distribution, respectively. For simplicity we suppress the argument  $(x_1, x_2)$ .

**Lemma 3.1.** *Set  $\delta := \delta(\mathbf{h}, u)$  as given in (4) and define for  $x_1, x_2 > 0$*

$$q_\psi^{(1)} := \frac{\log(x_2/x_1)}{2\sqrt{\delta}} + \sqrt{\delta} \quad q_\psi^{(2)} := \frac{\log(x_1/x_2)}{2\sqrt{\delta}} + \sqrt{\delta}, \quad (7)$$

$$V := \frac{1}{x_1} \Phi(q_\psi^{(1)}) + \frac{1}{x_2} \Phi(q_\psi^{(2)}). \quad (8)$$

*The partial derivatives of  $q_\psi^{(1)}$  and  $q_\psi^{(2)}$  are given by*

$$\frac{\partial q_\psi^{(1)}}{\partial x_1} = -\frac{1}{2\sqrt{\delta}x_1}, \quad \frac{\partial q_\psi^{(1)}}{\partial x_2} = \frac{1}{2\sqrt{\delta}x_2}, \quad \frac{\partial q_\psi^{(2)}}{\partial x_2} = -\frac{1}{2\sqrt{\delta}x_2}, \quad \frac{\partial q_\psi^{(2)}}{\partial x_1} = \frac{1}{2\sqrt{\delta}x_1}.$$

*The first and second order partial derivatives of  $V$  are given by*

$$\begin{aligned} \frac{\partial V}{\partial x_1} &= -\frac{1}{x_1^2} \Phi(q_\psi^{(1)}) - \frac{1}{2\sqrt{\delta}x_1^2} \varphi(q_\psi^{(1)}) + \frac{1}{2\sqrt{\delta}x_1x_2} \varphi(q_\psi^{(2)}), \\ \frac{\partial V}{\partial x_2} &= -\frac{1}{x_2^2} \Phi(q_\psi^{(2)}) - \frac{1}{2\sqrt{\delta}x_2^2} \varphi(q_\psi^{(2)}) + \frac{1}{2\sqrt{\delta}x_1x_2} \varphi(q_\psi^{(1)}), \\ \frac{\partial^2 V}{\partial x_1 \partial x_2} &= -\frac{2\sqrt{\delta} - q_\psi^{(1)}}{4\delta x_1^2 x_2} \varphi(q_\psi^{(1)}) - \frac{2\sqrt{\delta} - q_\psi^{(2)}}{4\delta x_1 x_2^2} \varphi(q_\psi^{(2)}). \end{aligned}$$

*Finally, the bivariate log-density is*

$$\log f(x_1, x_2) = -V + \log \left[ \left( \frac{\partial V}{\partial x_1} \right) \left( \frac{\partial V}{\partial x_2} \right) - \frac{\partial^2 V}{\partial x_1 \partial x_2} \right]. \quad (9)$$

The resulting parameter vector is  $\psi = (\theta_1, \alpha_1, \theta_2, \alpha_2)$ . We first define the pairwise likelihood for a general setting with  $M$  locations  $\mathbf{s}_1, \dots, \mathbf{s}_M$  and  $T$  time points  $0 \leq t_1 < \dots < t_T < \infty$ . In a second step we assume that the locations lie on a regular grid and that the time points are equidistant.

$$PL^{(M,T)}(\psi) = \sum_{i=1}^M \sum_{j=i+1}^M \sum_{k=1}^{T-1} \sum_{l=k+1}^T w_{i,j}^{(M)} w_{k,l}^{(T)} \log f_\psi(\eta(\mathbf{s}_i, t_k), \eta(\mathbf{s}_j, t_l)), \quad (10)$$

where  $w_{i,j}^{(M)} \geq 0$  and  $w_{k,l}^{(T)} \geq 0$  denote spatial and temporal weights, respectively. Since it is expected that space-time pairs, which are far apart in space or in time, have only little influence on the dependence parameters to be estimated, we define the weights, such that in the estimation only pairs with a maximal spatio-temporal distance of  $(r, p)$  are included, i.e.,

$$w_{i,j}^{(M)} = \mathbb{1}_{\{\|\mathbf{s}_i - \mathbf{s}_j\| \leq r\}}, \quad w_{k,l}^{(T)} = \mathbb{1}_{\{|t_k - t_l| \leq p\}}, \quad (11)$$

where  $\|\cdot\|$  denotes any arbitrary norm on  $\mathbb{R}^d$ . The pairwise likelihood estimates are given by

$$(\hat{\theta}_1, \hat{\alpha}_1, \hat{\theta}_2, \hat{\alpha}_2) = \arg \max_{(\theta_1, \alpha_1, \theta_2, \alpha_2)} PL^{(M,T)}(\theta_1, \alpha_1, \theta_2, \alpha_2). \quad (12)$$

Using the definition of the weights in (11), the log-likelihood function in (10) can be rewritten as

$$PL^{(M,T)}(\boldsymbol{\psi}) = \sum_{i=1}^M \sum_{\substack{j=i+1 \\ \|\mathbf{s}_i - \mathbf{s}_j\| \leq r}}^M \sum_{k=1}^{T-p} \sum_{l=k+1}^{\min\{k+p, T\}} \log f_{\boldsymbol{\psi}}(\eta(\mathbf{s}_i, t_k), \eta(\mathbf{s}_j, t_l)).$$

The following sampling scheme is assumed throughout.

**Condition 3.2.** We assume that the locations lie on a regular  $d$ -dimensional lattice,

$$S = \{\mathbf{s}_j, j = 1, \dots, m^d\} = \{\mathbf{s}_{(i_1, \dots, i_d)} = (i_1, \dots, i_d), i_1, \dots, i_d \in \{1, \dots, m\}\}.$$

Further assume that the time points are equidistant,

$$0 = t_0 \leq t_1 < \dots < t_T < \infty, \quad |t_k - t_{k-1}| = 1, \quad k = 1, \dots, T.$$

For later purposes, we rewrite the pairwise log-likelihood function under Condition 3.2 in the following way. Define  $\mathcal{H}_r$  as the set of all vectors with non-negative integer-valued components  $\mathbf{h}$  without the  $\mathbf{0}$ -vector, which point to other sites in the set of locations within distance  $r$ . Nott and Ryden [17] call this the *design mask*. We denote by  $|\mathcal{H}_r|$  the cardinality of the set  $\mathcal{H}_r$ . In our application, we will use the following design masks according to the Euclidean distance;

$$\begin{aligned} \mathcal{H}_1 &= \{(1, 0), (0, 1)\} \\ \mathcal{H}_2 &= \mathcal{H}_1 \cup \{(1, 1), (0, 2), (2, 0)\} \\ \mathcal{H}_3 &= \mathcal{H}_2 \cup \{(1, 2), (2, 1), (2, 2), (0, 3), (3, 0)\} \\ \mathcal{H}_4 &= \mathcal{H}_3 \cup \{(1, 3), (3, 1), (2, 3), (3, 2), (4, 0), (0, 4)\} \\ &\vdots \end{aligned}$$

Using Condition 3.2 and the design mask, the *pairwise log-likelihood function* in (12) can be rewritten as

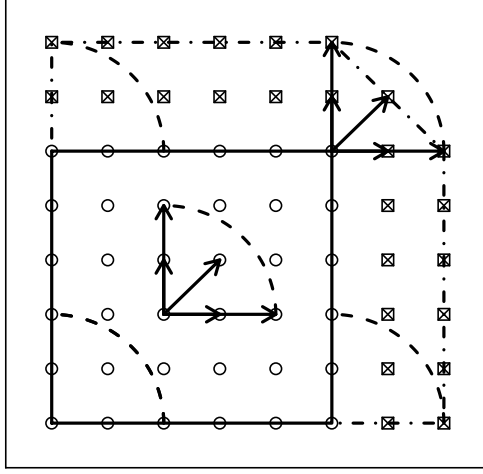
$$\begin{aligned} & PL^{(m^d, T)}(\boldsymbol{\psi}) \\ &= \sum_{i_1=1}^m \dots \sum_{i_d=1}^m \sum_{\substack{\mathbf{h} \in \mathcal{H}_r \\ \mathbf{s}_{(i_1, \dots, i_d)} + \mathbf{h} \in S}} \sum_{k=1}^T \sum_{l=k+1}^{\min\{k+p, T\}} \log f_{\boldsymbol{\psi}}(\eta(\mathbf{s}_{(i_1, \dots, i_d)}, t_k), \eta(\mathbf{s}_{(i_1, \dots, i_d)} + \mathbf{h}, t_l)) \\ &= \sum_{i_1=1}^m \dots \sum_{i_d=1}^m \sum_{k=1}^T g_{\boldsymbol{\psi}}(i_1, \dots, i_d, k; \mathcal{H}_r, p) - \mathcal{R}^{(m^d, T)}(\boldsymbol{\psi}), \end{aligned} \quad (13)$$

where

$$g_{\psi}(i_1, \dots, i_d, k; \mathcal{H}_r, p) = \sum_{\mathbf{h} \in \mathcal{H}_r} \sum_{l=k+1}^{k+p} \log f_{\psi}(\eta(\mathbf{s}_{(i_1, \dots, i_d)}, t_k), \eta(\mathbf{s}_{(i_1, \dots, i_d)} + \mathbf{h}, t_l)), \quad (14)$$

and  $\mathcal{R}^{(m^d, T)}(\psi)$  is a boundary term, given by

$$\mathcal{R}^{(m^d, T)}(\psi) = \sum_{i_1=1}^m \cdots \sum_{i_d=1}^m \sum_{\substack{\mathbf{h} \in \mathcal{H}_r \\ \mathbf{s}_{(i_1, \dots, i_d)} + \mathbf{h} \notin S}} \sum_{k=1}^T \sum_{\substack{l=k+1 \\ l > T}}^{k+p} \log f_{\psi}(\eta(\mathbf{s}_{(i_1, \dots, i_d)}, t_k), \eta(\mathbf{s}_{(i_1, \dots, i_d)} + \mathbf{h}, t_l)) \quad (15)$$



**Figure 3.1:** Visualization of the boundary term  $\mathcal{R}^{(m^d, T)}$  for  $d = 2$ ,  $m = 6$  and any time point, where the set  $S$  of locations is the inner square and the outer polygon represents the endpoints of pairs in the boundary

Figure 3.1 depicts a spatial grid with length  $m = 6$ , where the inner square is the set of observed locations  $S$  and the points in the outer polygon are endpoints of pairs which are in the boundary term  $\mathcal{R}^{(m^d, T)}$ . The figure visualizes the case  $\mathcal{H}_2$  which is represented by the quarter circles.

## 4 Strong consistency of the pairwise likelihood estimates

In this section we establish strong consistency for the pairwise likelihood estimates introduced in Section 3.2. For univariate time series models Davis and Yau [7] proved strong consistency of the composite likelihood estimates in full detail. For max-stable random fields with replicates, which are independent in time, Padoan et al. [18] showed consistency and asymptotic normality for the pairwise likelihood estimates. In contrast to previous studies, where either the spatial or the time domain increases, we show strong consistency as the space-time domain increases jointly.

#### 4.1 Ergodic properties for max-stable processes

Stoev and Taqqu [21] introduced extremal integrals as an analogy to sum-stable integrals. Based on the extremal integral representation of max-stable processes Stoev [20] establishes conditions under which the max-stable process is ergodic. Wang et al. [26] extend these results to a spatial setting. In the following, let  $\tau_{(h_1, \dots, h_d, u)}$  denote the multiparameter shift-operator. In accordance with the definitions and results in Wang et al. [26], we define ergodic and mixing space-time processes.

**Definition 4.1.** *A strictly stationary space-time process  $\{\eta(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in [0, \infty)\}$  is called ergodic, if for all  $A, B \in \sigma\{\eta(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in [0, \infty)\}$*

$$\lim_{m_1 \dots m_d T \rightarrow \infty} \frac{1}{m_1 \dots m_d T} \sum_{h_1=1}^{m_1} \dots \sum_{h_d=1}^{m_d} \sum_{u=1}^T P(A \cap \tau_{(h_1, \dots, h_d, u)}(B)) = P(A)P(B). \quad (16)$$

If the process satisfies additionally

$$\lim_{n \rightarrow \infty} P(A \cap \tau_{(s_{1,n}, \dots, s_{d,n}, t_n)}(B)) = P(A)P(B), \quad (17)$$

for all sequences  $\{(s_{1,n}, \dots, s_{d,n}, t_n), n \in \mathbb{N}\}$  with  $\max\{|s_{1,n}|, \dots, |s_{d,n}|, |t_n|\} \rightarrow \infty$ , we call the process mixing.

Note in (16) that in contrast to the ergodic theorem in Wang et al. [26], the number of terms in each sum is not equal, since we have an additional sum for the time component. We focus on max-stable processes with extremal integral representation

$$\eta(s_1, \dots, s_d, t) = \int_E^e U_{(s_1, \dots, s_d, t)}(f) dM_1, \quad (18)$$

where  $U_{(s_1, \dots, s_d, t)} : L^1(\mu) \rightarrow L^1(\mu)$  given by  $U_{(s_1, \dots, s_d, t)}(f) = f \circ \tau_{(s_1, \dots, s_d, t)}$  is a group of max-linear automorphisms with  $U_{(0, \dots, 0, 0)}(f) = f$  and the control measure  $\mu$  is the distribution of the space-time process,

$$\mu(A) = P(\eta(s_1, \dots, s_d, t) \in A), \quad A \in \sigma\{\eta(\mathbf{s}, t), (\mathbf{s}, t) \in \mathbb{R}^d \times [0, \infty)\}.$$

The following result is a direct extension of the uniparameter theorem established in Stoev [20], Theorem 3.4, and the multiparameter counterpart in Wang et al. [26].

**Proposition 4.2** (Wang et al. [26], Theorem 5.6). *The max-stable process defined in (18) is mixing, if and only if*

$$\int_E U_{(s_{1,n}, \dots, s_{d,n}, t_n)}(f) \wedge U_{(0, \dots, 0, 0)}(f) d\mu = \int_E U_{(s_{1,n}, \dots, s_{d,n}, t_n)}(f) \wedge f d\mu \rightarrow 0, \quad (19)$$

for all sequences  $\{(s_{1,n}, \dots, s_{d,n}, t_n)\}$  with  $\max\{|s_{1,n}|, \dots, |s_{d,n}|, |t_n|\} \rightarrow \infty$  as  $n \rightarrow \infty$ .

Wang et al. [26] showed, that the ergodic theorem stated above holds for mixing max-stable processes with extremal integral representation (18) in the case of  $T = m$ . The extension to the multiparameter case where  $T \neq m$  is a simple generalisation using Theorem 6.1.2 in Krengel [15], which is a multiparameter extension of the Akcoglus ergodic theorem. Ergodic properties of Brown-Resnick processes have been studied for the uniparameter case in Stoev and Taqqu [21] and Wang and Stoev [27]. We summarize the results in the following proposition.



**Proposition 4.3.** *The Brown-Resnick process in Proposition 2.2 with extremal integral representation*

$$\left\{ \int_E^e \exp \{W(\mathbf{s}, t) - \delta(\mathbf{s}, t)\} dM_1 \quad \mathbf{s} \in \mathbb{R}^d, t \in [0, \infty) \right\}$$

*is mixing in space and time. The strong law of large numbers holds;*

$$\frac{1}{m^d T} \sum_{i_1=1}^m \cdots \sum_{i_d=1}^m \sum_{k=1}^T g(\eta(\mathbf{s}_{(i_1, \dots, i_d)}, t_k)) \xrightarrow{a.s.} \mathbb{E} [g(\eta(\mathbf{s}_{(1, \dots, 1)}, t_1))], \quad mT \rightarrow \infty \quad (20)$$

where  $g$  is a measurable function.

## 4.2 Consistency for large $mT$

In the following we show that the pairwise likelihood estimate resulting from maximizing (13) for the model defined in Proposition 2.2 is strongly consistent.

**Theorem 4.4.** *Assume that the true parameter vector  $\boldsymbol{\psi}^* = (\theta_1^*, \alpha_1^*, \theta_2^*, \alpha_2^*)$  lies in a compact set  $\Psi$ , which does not contain  $\mathbf{0}$  and which satisfies for some  $c > 0$*

$$\Psi \subseteq \{\min \{\theta_1, \theta_2\} > c, \alpha_1, \alpha_2 \in (0, 2]\}. \quad (21)$$

*Assume also that the identifiability condition*

$$\boldsymbol{\psi} = \tilde{\boldsymbol{\psi}} \quad \Leftrightarrow \quad f_{\boldsymbol{\psi}}(\eta(\mathbf{s}_1, t_1), \eta(\mathbf{s}_2, t_2)) = f_{\tilde{\boldsymbol{\psi}}}(\eta(\mathbf{s}_1, t_1), \eta(\mathbf{s}_2, t_2)), \quad a.s. \quad (22)$$

*is satisfied for all  $(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)$ . It then follows that the pairwise likelihood estimate*

$$\hat{\boldsymbol{\psi}}_{m^d, T} = \arg \max_{\boldsymbol{\psi} \in \Psi} PL^{(m^d, T)}(\boldsymbol{\psi})$$

*is strongly consistent, i.e.  $\hat{\boldsymbol{\psi}}_{m^d, T} \xrightarrow{a.s.} \boldsymbol{\psi}^*$  as  $mT \rightarrow \infty$ .*

**Remark 4.5.** For the identifiability assumption (22) we consider different cases according to the maximal space-time lag  $(r, p)$  included in the composite likelihood. Recall that the pairwise density, see Lemma 3.1, depends on the spatial distance  $\mathbf{h}$  and the time lag  $u$  only through the function  $\delta(\mathbf{h}, u) = \theta_1 \|\mathbf{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2}$ . For specific combinations of  $(r, p)$  not all parameters are identifiable. For example, if the maximal spatial lag taken into account in the estimation equals one, i.e.  $r = 1$ , and  $p > 1$ , the parameter  $\alpha_1$  is not identifiable. Strong consistency still holds for the remaining parameters. Table 1 lists the various scenarios.

*Proof of Theorem 4.4.* To show strong consistency of the estimates we follow the method of Wald [25]. Accordingly, it suffices to show the following conditions.

(C1) Strong law of large numbers: Uniformly on the compact set  $\Psi$ ,

$$\frac{1}{m^d T} PL^{(m^d, T)}(\boldsymbol{\psi}) \xrightarrow{a.s.} PL(\boldsymbol{\psi}) := \mathbb{E} [g_{\boldsymbol{\psi}}(1, \dots, 1, 1; \mathcal{H}_r, p)], \quad mT \rightarrow \infty.$$

(C2) The function  $PL(\boldsymbol{\psi})$  is uniquely maximized at the true parameter vector  $\boldsymbol{\psi}^* \in \Psi$ .

Maximal spatial lag $r$	Maximal temporal lag $p$	Identifiable parameters
0	1	$\theta_2$
0	$p, p > 1$	$\theta_2, \alpha_2$
1	0	$\theta_1$
$r, r > 1$	0	$\theta_1, \alpha_1$
1	1	$\theta_1, \theta_2$
1	$p, p > 1$	$\theta_1, \theta_2, \alpha_2$
$r, r > 1$	1	$\theta_1, \alpha_1, \theta_2$
$r, r > 1$	$p, p > 1$	$\theta_1, \alpha_1, \theta_2, \alpha_2$

**Table 1:** Identifiable parameters corresponding to different maximal space-time lags  $(r, p)$  included in the pairwise likelihood function.

From (C1) and (C2) strong consistency follows. First we prove (C1). Recall from (13) that the pairwise likelihood function is given by

$$PL^{(m^d, T)}(\boldsymbol{\psi}) = \sum_{i_1=1}^m \cdots \sum_{i_d=1}^m \sum_{k=1}^T g_{\boldsymbol{\psi}}(i_1, \dots, i_d, k; \mathcal{H}_r, p) - \mathcal{R}^{(m^d, T)}(\boldsymbol{\psi}),$$

where  $g_{\boldsymbol{\psi}}$  and  $\mathcal{R}^{(m^d, T)}(\boldsymbol{\psi})$  are defined in (14) and (15), respectively. The pointwise convergence of the first term on the right hand side to  $PL(\boldsymbol{\psi})$  follows immediately from Proposition 4.3 together with the fact that  $g_{\boldsymbol{\psi}}$  in (14) is a measurable function of lagged versions of  $\eta(\mathbf{s}_{(i_1, \dots, i_d), t_k})$ .

In the following, we show that the convergence is uniform and that the boundary term defined in (15) converges to zero almost surely. For both steps, observe first that we can bound the log-density from Lemma 3.1. For  $x_1, x_2 > 0$

$$\begin{aligned} |\log f_{\boldsymbol{\psi}}(x_1, x_2)| &= \left| -V + \log \left( \frac{\partial V}{\partial x_1} \frac{\partial V}{\partial x_2} - \frac{\partial^2 V}{\partial x_1 \partial x_2} \right) \right| \\ &\leq \left| -\frac{1}{x_1} \Phi(q_{\boldsymbol{\psi}}^{(1)}) \right| + \left| -\frac{1}{x_1} \Phi(q_{\boldsymbol{\psi}}^{(2)}) \right| + \left| \frac{\partial V}{\partial x_1} \frac{\partial V}{\partial x_2} - \frac{\partial^2 V}{\partial x_1 \partial x_2} \right| \\ &\leq \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_1^2 x_2^2} + \frac{1}{2\sqrt{\delta(\mathbf{h}, u)}} \left( \frac{1}{x_1^2 x_2^2} + \frac{1}{x_1^3 x_2} + \frac{1}{x_1^2 x_2^2} + \frac{1}{x_1 x_2^3} + \frac{1}{x_1^2 x_2} + \frac{1}{x_1 x_2^2} \right) \\ &\quad + \frac{1}{4\delta(\mathbf{h}, u)} \left( \frac{1}{x_1^2 x_2^2} + \frac{1}{x_1^3 x_2} + \frac{1}{x_1 x_2^3} + \frac{1}{x_1^2 x_2^2} + \left| \frac{q_{\boldsymbol{\psi}}^{(1)}}{x_1^2 x_2} + \frac{q_{\boldsymbol{\psi}}^{(2)}}{x_1 x_2^2} \right| \right), \end{aligned}$$

where  $q_{\boldsymbol{\psi}}^{(1)}, q_{\boldsymbol{\psi}}^{(2)}$  and  $V$  are defined in (7) and (8), respectively, where  $\Phi(\cdot) \leq 1$  and  $\varphi(\cdot) \leq 1$  were used. Since the marginal distributions of the max-stable space-time process are assumed to be standard Fréchet, it follows that for every fixed location  $\mathbf{s} \in S$  and fixed time point  $t \in T$   $1/\eta(\mathbf{s}, t)$  is standard exponentially distributed. Using Hölder's inequality, it follows that

$$\mathbb{E}_{\boldsymbol{\psi}^*} [|\log f_{\boldsymbol{\psi}}(\eta(\mathbf{s}_1, t_1), \eta(\mathbf{s}_2, t_2))|] \leq K_1 + \frac{K_2}{2\sqrt{\delta(\mathbf{h}, u)}} + \frac{K_3}{4\delta(\mathbf{h}, u)},$$

where  $K_1, K_2, K_3 > 0$  are finite constants. Since the parameter space  $\Psi$  is assumed to be compact and together with assumption (21),  $\delta$  can be bounded away from zero, i.e.

$$\begin{aligned} \delta(\mathbf{h}, u) &= \theta_1 \|\mathbf{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2} \geq \min\{\theta_1, \theta_2\} (\|\mathbf{h}\|^{\alpha_1} + |u|^{\alpha_2}) \\ &> c(\|\mathbf{h}\|^{\alpha_1} + |u|^{\alpha_2}) > \tilde{c} > 0, \end{aligned} \tag{23}$$

since  $\alpha_1, \alpha_2 \in (0, 2]$ , where  $\tilde{c} > 0$  is some constant independent of the parameters. Therefore,

$$\mathbb{E}_{\psi^*} [|\log f_{\psi}(\eta(\mathbf{s}_1, t_1), \eta(\mathbf{s}_2, t_2))|] < K_1 + \frac{K_2}{2\sqrt{\tilde{c}}} + \frac{K_3}{4\tilde{c}} =: K_4 < \infty, \quad (24)$$

where  $K_4 > 0$ . Note that in the same way we can show that the expectation of the squared bivariate log-density is finite, since it only involves higher order moments of the exponential distribution.

To establish uniform convergence, we follow Straumann and Mikosch [22], Theorem 2.7, and show that

$$\mathbb{E} \left[ \sup_{\psi \in \Psi} |g_{\psi}(1, \dots, 1, 1; \mathcal{H}_r, p)| \right] < \infty.$$

It is sufficient to verify that

$$\mathbb{E} \left[ \sup_{\psi \in \Psi} |\log f_{\psi}(\eta(\mathbf{s}_{(i_1, \dots, i_d)}, t_k), \eta(\mathbf{s}_{(i_1, \dots, i_d)} + \mathbf{h}, t_k + u))| \right] < \infty.$$

Since the pairwise density is continuous and because of the compact parameter space, the statement follows immediately using (23) and (24).

As a last step for (C1) we show that the boundary term  $\mathcal{R}^{(m^d, T)}(\psi)$  converges to 0 almost surely. For notational simplicity, we only consider the case  $d = 2$ . The general case is proved analogously. First note from (15) that

$$\begin{aligned} & \mathbb{E} \left[ \left| \frac{1}{m^2 T} \mathcal{R}^{(m^2, T)}(\psi) \right| \right] \\ & \leq \frac{1}{m^2 T} \sum_{i_1=1}^m \sum_{i_2=1}^m \sum_{\substack{\mathbf{h} \in \mathcal{H}_r \\ \mathbf{s}_{(i_1, i_2)} + \mathbf{h} \notin S}} \sum_{k=1}^T \sum_{\substack{l=k+1 \\ l > T}}^{k+p} \mathbb{E} [|\log f_{\psi}(\eta(\mathbf{s}_{(i_1, i_2)}, t_k), \eta(\mathbf{s}_{(i_1, i_2)} + \mathbf{h}, t_l))|] \\ & \leq \frac{1}{m^2 T} \sum_{i_1=1}^m \sum_{i_2=1}^m \sum_{\substack{\mathbf{h} \in \mathcal{H}_r \\ \mathbf{s}_{(i_1, i_2)} + \mathbf{h} \notin S}} \sum_{k=1}^T \sum_{\substack{l=k+1 \\ l > T}}^{k+p} K_4 \leq \frac{K_4 K_5}{mT} \rightarrow 0, \quad mT \rightarrow \infty, \end{aligned}$$

where we used the bound derived in (24) and the fact that the number of space-time points in the boundary is of order  $m$  (independent of  $T$ ) and, therefore, can be bounded by  $K_5 m$  with  $K_5 > 0$  a constant independent of  $m$  and  $T$ . We write  $\mathcal{R}^{(m^2, T)}(\psi)$  in the following way using the function  $g_{\psi}$  in (14). Denote by  $\mathcal{B}_{m, T}$  the set of space-time indices  $(i_1, i_2, k)$  for which  $\mathbf{s}_{(i_1, i_2)} + \mathbf{h} \notin S$  or  $k > T$ . In Figure 3.1  $\mathcal{B}_{m, T}$  corresponds to the indices of the locations  $\mathbf{s}_{(i_1, i_2)} \in S$  for which  $\mathbf{s}_{(i_1, i_2)} + \mathbf{h}$  is in the outer polygon. The cardinality of the set  $\mathcal{B}_{m, T}$  can be bounded by using the maximum norm instead of the euclidean norm, i.e.

$$|\mathcal{B}_{m, T}| \leq r(2m + 1) =: K_5 m.$$

The boundary term in (15) can then be written as

$$\mathcal{R}^{(m^2, T)}(\psi) = \sum_{(i_1, i_2, k) \in \mathcal{B}_{m, T}} g_{\psi}(i_1, i_2, k; \mathcal{H}_r, p),$$

where  $g_{\psi}$  is defined in (14). In the same way as before, it follows by the strong law of large numbers that

$$\frac{1}{|\mathcal{B}_{m, T}|} \sum_{(i_1, i_2, k) \in \mathcal{B}_{m, T}} g_{\psi}(i_1, i_2, k; \mathcal{H}_r, p) \xrightarrow{a.s.} \mathbb{E} [g_{\psi}(1, 1, 1; \mathcal{H}_r, p)],$$

uniformly on the compact set  $\Psi$ . Therefore,

$$\frac{1}{m^2 T} \mathcal{R}^{(m^2, T)}(\psi) \leq \frac{K_5}{mT} \frac{1}{|\mathcal{B}_{m, T}|} \sum_{(i_1, i_2, k) \in \mathcal{B}_{m, T}} g_\psi(i_1, i_2, k; \mathcal{H}_r, p) \xrightarrow{a.s.} 0,$$

since  $\mathbb{E}[|g_\psi(1, 1, 1; \mathcal{H}_r, p)|] < \infty$ . This proves (C1).

To prove (C2), note by Jensen's inequality that

$$\mathbb{E}_{\psi^*} \left[ \log \left( \frac{f_\psi(x_1, x_2)}{f_{\psi^*}(x_1, x_2)} \right) \right] \leq \log \left( \mathbb{E}_{\psi^*} \left[ \frac{f_\psi(x_1, x_2)}{f_{\psi^*}(x_1, x_2)} \right] \right) = 0$$

and, hence,

$$PL(\psi) \leq PL(\psi^*).$$

So,  $\psi^*$  maximizes  $PL(\psi)$  and is the unique optimum if and only if there is equality in Jensen's inequality. However, this is precluded by (22).  $\square$

## 5 Asymptotic normality of the pairwise likelihood estimates

In order to prove asymptotic normality of the pairwise likelihood estimates resulting from maximizing (13) we need the following results for the pairwise log-density. The proofs can be found in Appendix A.

**Lemma 5.1.** (1) *The gradient of the bivariate log-density satisfies*

$$\mathbb{E}_{\psi^*} \left[ |\nabla_\psi \log f_\psi(\eta(\mathbf{s}_1, t_1), \eta(\mathbf{s}_2, t_2))|^3 \right] < \infty$$

(2) *The Hessian of the pairwise log-density satisfies*

$$\mathbb{E}_{\psi^*} \left[ \sup_{\psi \in \Psi} |\nabla_\psi^2 \log f_\psi(\eta(\mathbf{s}_1, t_1), \eta(\mathbf{s}_2, t_2))| \right] < \infty.$$

Assuming asymptotic normality of the pairwise score function, then it is relatively routine to show that the pairwise likelihood estimates are asymptotically normal. To formulate the result, recall from (13) that the pairwise likelihood function can be written as

$$PL^{(m^d, T)}(\psi) = \sum_{i_1=1}^m \cdots \sum_{i_d=1}^m \sum_{k=1}^T g_\psi(i_1, \dots, i_d, k; \mathcal{H}_r, p) - \mathcal{R}^{(m^d, T)}(\psi),$$

where  $g_\psi$  is defined in (14). The pairwise score function is then given by

$$\sum_{i_1=1}^m \cdots \sum_{i_d=1}^m \sum_{k=1}^T \nabla_\psi g_\psi(i_1, \dots, i_d, k; \mathcal{H}_r, p) - \nabla_\psi \mathcal{R}^{(m^d, T)}(\psi),$$

where  $\nabla_\psi g_\psi(i_1, \dots, i_d, k; \mathcal{H}_r, p)$  is the gradient of the function  $g_\psi$  with respect to  $\psi$ .

**Theorem 5.2.** *Assume that the conditions of Theorem 4.4 hold. In addition, assume that a central limit theorem holds for  $\nabla_\psi g_\psi(i_1, \dots, i_d, k; \mathcal{H}_r, p)$  in the following sense*

$$\frac{1}{m^{d/2} \sqrt{T}} \sum_{i_1=1}^m \cdots \sum_{i_d=1}^m \sum_{k=1}^T \nabla_\psi g_{\psi^*}(i_1, \dots, i_d, k; \mathcal{H}_r, p) \xrightarrow{d} \mathcal{N}(0, \Sigma), \quad mT \rightarrow \infty, \quad (25)$$

where  $\psi^*$  is the true parameter vector and  $\Sigma$  is some covariance matrix. Then it follows that the pairwise likelihood estimates  $\hat{\psi}_{m^d, T}$  satisfy

$$m^{d/2} \sqrt{T} (\hat{\psi}_{m^d, T} - \psi^*) \xrightarrow{d} \mathcal{N}(0, F^{-1} \Sigma (F^{-1})^T), \quad mT \rightarrow \infty,$$

where

$$F = \mathbb{E}_{\psi^*} [-\nabla_{\psi}^2 g_{\psi^*}(1, \dots, 1, 1; \mathcal{H}_r, p)].$$

*Proof.* We use a standard Taylor expansion of the pairwise score function around the true parameter vector:

$$\begin{aligned} m^{d/2} \sqrt{T} (\hat{\psi}_{m^d, T} - \psi^*) &= - \left( \frac{1}{m^d T} \nabla_{\psi}^2 PL^{(m^d, T)}(\tilde{\psi}) \right)^{-1} \left( \frac{1}{m^{d/2} \sqrt{T}} \nabla_{\psi} PL^{(m^d, T)}(\psi^*) \right) \\ &= - \left( \frac{1}{m^d T} \sum_{i_1=1}^m \cdots \sum_{i_d=1}^m \sum_{k=1}^T \nabla_{\psi}^2 g_{\tilde{\psi}}(i_1, \dots, i_d, k; \mathcal{H}_r, p) - \nabla_{\psi}^2 \mathcal{R}^{(m^d, T)}(\tilde{\psi}) \right)^{-1} \\ &\quad \times \left( \frac{1}{m^{d/2} \sqrt{T}} \sum_{i_1=1}^m \cdots \sum_{i_d=1}^m \sum_{k=1}^T \nabla_{\psi} g_{\psi^*}(i_1, \dots, i_d, k; \mathcal{H}_r, p) - \nabla_{\psi} \mathcal{R}^{(m^d, T)}(\psi^*) \right), \end{aligned}$$

where  $\tilde{\psi} \in [\hat{\psi}_{m^d, T}, \psi^*]$ . For now, we ignore the boundary term and analyze the first terms in the outer brackets. By (25) the second term converges to a normal distribution with mean 0 and covariance matrix  $\Sigma$ . For the first part we use the same arguments as in the consistency proof and show a strong law of large numbers. Since the underlying space-time process in the likelihood function is mixing, it follows that the process

$\left\{ \nabla_{\psi}^2 g_{\psi}(i_1, \dots, i_d, k; \mathcal{H}_r, p), \mathbf{s}_{(i_1, \dots, i_d)} \in \mathbb{Z}^d, t_k \in \mathbb{Z} \right\}$  is mixing as a measurable function of mixing and lagged processes. To prove the uniform convergence we need to verify that

$$\mathbb{E}_{\psi^*} \left[ \sup_{\psi \in \Psi} |\nabla_{\psi}^2 g_{\psi}(1, \dots, 1, 1; \mathcal{H}_r, p)| \right] < \infty.$$

This follows immediately from Lemma 5.1. Putting this together with the fact that  $\tilde{\psi} \in [\hat{\psi}_{m^d, T}, \psi^*]$ , and because of the strong consistency, it follows that

$$\frac{1}{m^d T} \sum_{i_1=1}^m \cdots \sum_{i_d=1}^m \sum_{k=1}^T \nabla_{\psi}^2 g_{\tilde{\psi}}(i_1, \dots, i_d, k; \mathcal{H}_r, p) \xrightarrow{a.s.} \mathbb{E}_{\psi^*} [\nabla_{\psi}^2 g_{\psi^*}(1, \dots, 1, 1; \mathcal{H}_r, p)] =: -F.$$

For the boundary term  $\mathcal{R}^{(m^d, T)}$  observe that it can be written as

$$\mathcal{R}^{(m^d, T)}(\psi) = \sum_{(i_1, \dots, i_d, k) \in \mathcal{B}_{m, T}} g_{\psi}(i_1, \dots, i_d, k; \mathcal{H}_r, p).$$

Using assumption (25) together with the strong law of large numbers for  $\left\{ \nabla_{\psi}^2 g_{\psi}(i_1, \dots, i_d, k; \mathcal{H}_r, p) \right\}$  it follows in the same way as in the proof of Theorem 4.4 that

$$\frac{1}{m^d T} \nabla_{\psi}^2 \mathcal{R}^{(m^d, T)} \xrightarrow{a.s.} 0, \quad \text{and} \quad \frac{1}{m^{d/2} \sqrt{T}} \nabla_{\psi} \mathcal{R}^{(m^d, T)} \xrightarrow{P} 0.$$

Combining these results, we obtain

$$m^{d/2} \sqrt{T} (\hat{\psi}_{m^d, T} - \psi^*) \xrightarrow{d} \mathcal{N}(0, F^{-1} \Sigma (F^{-1})^T), \quad mT \rightarrow \infty.$$

□

In the next section we provide a sufficient condition for (25).

## 5.1 Asymptotic normality and $\alpha$ -mixing

In this section we consider asymptotic normality of the parameters estimates for the process in Proposition 2.2. Under the assumption of  $\alpha$ -mixing on the random field the key is to show asymptotic normality for the score function of the pairwise likelihood. For an increasing time domain and fixed number of locations asymptotic normality of the pairwise likelihood estimates was shown in Huser and Davison [11]. The main difference between a temporal setting and a space-time setting is the definition of the  $\alpha$ -mixing coefficients and the resulting assumptions to obtain a central limit theorem for the score function.

We apply the central limit theorem for random fields established in Bolthausen [2] to the score function of the pairwise likelihood in our model. In a second step we verify the  $\alpha$ -mixing conditions for the max-stable process introduced in Section 2. First, we define the  $\alpha$ -mixing coefficients in a space-time setting as follows. Define the distances

$$d((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \max \left\{ \max_{1 \leq i \leq d} |\mathbf{s}_1(i) - \mathbf{s}_2(i)|, |t_1 - t_2| \right\}, \quad \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{Z}^d, t_1, t_2 \in \mathbb{N}$$

$$d(\Lambda_1, \Lambda_2) = \inf \{d((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)), (\mathbf{s}_1, t_1) \in \Lambda_1, (\mathbf{s}_2, t_2) \in \Lambda_2\}, \quad \Lambda_1, \Lambda_2 \subset \mathbb{Z}^d \times \mathbb{N}.$$

Let further  $\mathcal{F}_{\Lambda_i} = \sigma \{ \eta(\mathbf{s}, t), (\mathbf{s}, t) \in \Lambda_i \}$  for  $i = 1, 2$ . The mixing coefficients are defined for  $k, l, n \geq 0$  by

$$\alpha_{k,l}(n) = \sup \{ |P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}_{\Lambda_i}, |\Lambda_1| \leq k, |\Lambda_2| \leq l, d(\Lambda_1, \Lambda_2) \geq n \} \quad (26)$$

and depend on the sizes and the distance of the sets  $\Lambda_1$  and  $\Lambda_2$ . A space-time process is called  $\alpha$ -mixing, if  $\alpha_{k,l}(n) \rightarrow 0$  as  $n \rightarrow \infty$  for all  $k, l \geq 0$ . We assume that the process  $\{ \eta(\mathbf{s}, t), (\mathbf{s}, t) \in \mathbb{Z}^d \times \mathbb{N} \}$  is  $\alpha$ -mixing with mixing coefficients defined in (26), from which it follows that the score process

$$\left\{ \nabla_{\psi} g_{\psi}(i_1, \dots, i_d, k; \mathcal{H}_r, p), (\mathbf{s}_{(i_1, \dots, i_d)}, t_k) \in \mathbb{Z}^d \times \mathbb{N} \right\}. \quad (27)$$

is  $\alpha$ -mixing. We apply Bolthausen's central limit theorem to the process in (27). By adjusting the assumptions on the  $\alpha$ -mixing coefficients to the score process, we obtain the following proposition.

**Proposition 5.3.** *Assume, that the following conditions hold:*

- (1) *The process  $\{ (\eta(\mathbf{s}, t), (\mathbf{s}, t) \in \mathbb{Z}^d \times \mathbb{N}) \}$  is strongly mixing with mixing coefficients  $\alpha_{k,l}(n)$  as in (26).*
- (2)  *$\sum_{n=1}^{\infty} n^d \alpha_{k,l}(n) < \infty$  for  $k + l \leq 4(|\mathcal{H}_r| + 1)(p + 1)$  and  $\alpha_{(|\mathcal{H}_r|+1)(p+1), \infty}(n) = o(n^{-(d+1)})$ .*
- (3) *There exists  $\beta > 0$  such that*

$$\mathbb{E} \left[ |\nabla_{\psi} g_{\psi^*}((i_1, \dots, i_d), k, \mathcal{H}_r, p)|^{2+\beta} \right] < \infty \quad \text{and}$$

$$\sum_{n=1}^{\infty} n^d \alpha_{(|\mathcal{H}_r|+1)(p+1), (|\mathcal{H}_r|+1)(p+1)}(n)^{\beta/(2+\beta)} < \infty.$$

Then,

$$\frac{1}{m^{d/2} \sqrt{T}} \sum_{i_1=1}^m \cdots \sum_{i_d=1}^m \sum_{k=1}^T \nabla_{\psi} g_{\psi^*}(i_1, \dots, i_d, k; \mathcal{H}_r, 0) \xrightarrow{d} \mathcal{N}(0, \Sigma), \quad mT \rightarrow \infty,$$

where  $\Sigma = \sum_{\mathbf{s}_{(i_1, \dots, i_d)} \in \mathbb{Z}^d} \sum_{t_k \in \mathbb{N}} \text{Cov}(\nabla_{\psi} g_{\psi^*}(1, \dots, 1, 1; \mathcal{H}_r, p), \nabla_{\psi} g_{\psi^*}(i_1, \dots, i_d, k; \mathcal{H}_r, p)).$

In a second step, we want to analyze the strong mixing property and the related assumptions for our model. Recent work by Dombry and Eyi-Minko [9] deals with strong mixing properties for max-stable random fields. By using point process representation of max-stable processes together with coupling techniques, they showed that the  $\alpha$ -mixing coefficients can be bounded by a function of the tail dependence coefficient. A direct extension to the space-time setting gives the following lemma.

**Lemma 5.4** (Dombry and Eyi-Minko [9], Corollary 2.2). *Consider the stationary max-stable space-time process  $\{\eta(\mathbf{s}, t), (\mathbf{s}, t) \in \mathbb{Z}^d \times \mathbb{N}\}$  with tail dependence coefficient  $\chi(\mathbf{h}, u)$ . The  $\alpha$ -mixing coefficients in (26) satisfy*

$$\alpha_{k,l}(n) \leq kl \sup_{\max\{\|\mathbf{h}\|, |u|\} \geq n} \chi(\mathbf{h}, u) \quad \text{and} \quad \alpha_{k,\infty}(n) \leq k \sum_{\max\{\|\mathbf{h}\|, |u|\} \geq n} \chi(\mathbf{h}, u).$$

For the model described in Proposition 2.2 with tail dependence coefficient  $\chi$  in (5), it follows by using the inequality for the normal tail probability  $\bar{\Phi}(x) \leq e^{-x^2/2}$  that

$$\begin{aligned} \alpha_{k,l}(n) &\leq 4kl \sup_{\max\{\|\mathbf{h}\|, |u|\} \geq n} (1 - \Phi(\sqrt{\delta(\mathbf{h}, u)})) \leq 4kl \sup_{\max\{\|\mathbf{h}\|, |u|\} \geq n} \exp\left\{-\frac{\delta(\mathbf{h}, u)}{2}\right\} \\ &= 4kl \sup_{\max\{\|\mathbf{h}\|, |u|\} \geq n} \exp\left\{-\frac{1}{2}(\theta_1 \|\mathbf{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2})\right\} \\ &\leq 4kl \sup_{\max\{\|\mathbf{h}\|, |u|\} \geq n} \exp\left\{-\frac{1}{2} \min\{\theta_1, \theta_2\} (\max\{\|\mathbf{h}\|, |u|\})^{\min\{\alpha_1, \alpha_2\}}\right\}. \end{aligned}$$

For  $n \rightarrow \infty$ , this tends to zero for all  $k, l \geq 0$ . Thus,  $\{\eta(\mathbf{s}, t), (\mathbf{s}, t) \in \mathbb{Z}^d \times \mathbb{N}\}$  is strongly mixing. This shows the first assertion in Corollary 5.3. Furthermore, for  $k + l \leq 4(|\mathcal{H}_r| + 1)(p + 1)$  the coefficients satisfy

$$\begin{aligned} \sum_{n=1}^{\infty} n^d \alpha_{k,l}(n) &\leq 4kl \sum_{n=1}^{\infty} n^d \sup_{\max\{\|\mathbf{h}\|, |u|\} \geq n} \exp\left\{-\frac{1}{2}(\theta_1 \|\mathbf{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2})\right\} \\ &\leq 4kl \sum_{n=1}^{\infty} n^d \exp\left\{-\frac{1}{2} \min\{\theta_1, \theta_2\} n^{\min\{\alpha_1, \alpha_2\}}\right\} < \infty. \end{aligned}$$

In addition,

$$n^{d+1} \alpha_{(|\mathcal{H}_r|+1)(p+1), \infty}(n) \leq n^{d+1} (|\mathcal{H}_r| + 1)(p + 1) \sum_{x \geq n} \exp\left\{-\frac{1}{2} \min\{\theta_1, \theta_2\} x^{\min\{\alpha_1, \alpha_2\}}\right\},$$

which proves (2) in Proposition 5.3. As for (3), from Lemma 5.1 and using  $\beta = 1$  we know that

$$\mathbb{E} \left[ \left| \nabla_{\psi} g_{\psi^*}((i_1, \dots, i_d), k, \mathcal{H}_r, p) \right|^{(2+\beta)} \right] < \infty.$$

Using the same arguments as above, it is easy to see that the second condition in (3) holds. By combining the above results with Theorem 5.2 we obtain asymptotic normality for the parameter estimates  $\hat{\psi}_{m^d, T}$  for an increasing number of space-time locations.

**Theorem 5.5.** *Assume that the conditions of Theorem 4.4 hold. Then,*

$$(m^d T)^{1/2} (\hat{\psi}_{m^d, T} - \psi^*) \xrightarrow{d} \mathcal{N}(0, F^{-1} \Sigma (F^{-1})^\top), \quad mT \rightarrow \infty,$$

with

$$F = \mathbb{E}_{\psi^*} [-\nabla_{\psi}^2 g_{\psi^*}(1, \dots, 1, 1; \mathcal{H}_r, p)]$$

and

$$\Sigma = \sum_{\mathbf{s}_{(i_1, \dots, i_d)} \in \mathbb{Z}^d} \sum_{t_k \in \mathbb{N}} \mathbb{Cov}(\nabla_{\psi} g_{\psi^*}(1, \dots, 1, 1; \mathcal{H}_r, p), \nabla_{\psi} g_{\psi^*}(i_1, \dots, i_d, k; \mathcal{H}_r, p)).$$

## 6 Simulation study

We illustrate the small sample behaviour of the pairwise likelihood estimation for spatial dimension  $d = 2$  in a simulation experiment. The setup for this study is:

1. The spatial locations consisted of a  $10 \times 10$  grid

$$S = \{\mathbf{s}_{(i_1, i_2)} = (i_1, i_2), i_1, i_2 \in \{1, \dots, 10\}\}.$$

The time points are chosen equidistantly,  $1 < \dots T = 100$ .

2. One hundred independent Gaussian space-time processes  $Z_j(s_n \mathbf{s}, t_n t), j = 1, \dots, 100$  were generated using the R-package `RandomFields` with covariance function  $\rho(s_n \mathbf{h}, t_n u)$ . We use the following correlation function for the underlying Gaussian random field.

$$\rho(\mathbf{h}, u) = (1 + \theta_1 \|\mathbf{h}\|^{\alpha_1} + \theta_2 |u|^{\alpha_2})^{-3/2}.$$

Assumption 2.1 is fulfilled and the limit function  $\delta$  is given by

$$\lim_{n \rightarrow \infty} \log n(1 - \rho(s_n \mathbf{h}, t_n u)) = \delta(\mathbf{h}, u) = \frac{3}{2} \theta_1 \|\mathbf{h}\|^{\alpha_1} + \frac{3}{2} \theta_2 |u|^{\alpha_2}.$$

3. The simulated processes were transformed to standard Fréchet margins using the transformation  $-1/\log(\Phi(Z_j(\mathbf{s}, t)))$  for  $\mathbf{s} \in S$  and  $t \in \{t_1, \dots, t_T\}$ .
4. The pointwise maximum of the transformed Gaussian random fields was computed and rescaled by  $1/n$  to obtain an approximation of a max-stable random field, i.e.

$$\eta(\mathbf{s}, t) = \frac{1}{100} \bigvee_{j=1}^{100} -\frac{1}{\log(\Phi(Z_j(s_n \mathbf{s}, t_n t)))}, \quad \mathbf{s} \in S, t \in \{t_1, \dots, t_T\}.$$

5. The parameters  $\theta_1, \alpha_1, \theta_2$  and  $\alpha_2$  for different combinations of maximal space-time lags  $(r, p)$  were estimated by maximizing (13). The program is adjusted such that it takes care of identifiability issues, when some of the parameters are not identifiable, cf. Remark 4.5.
6. Steps 1. - 5. are repeated 100 times.

Figures 6.1 and 6.2 show the resulting estimates as a function of  $(r, p)$ , where the true parameter set is given by  $\psi^* = (\theta_1^*, \alpha_1^*, \theta_2^*, \alpha_2^*) = (0.06, 1, 0.04, 1)$ . Figure 6.1 shows the resulting estimates for the spatial parameters  $\theta_1$  and  $\alpha_1$ . The horizontal axis shows the different maximal space-time lags included in the pairwise likelihood function from (13). Each vertical dot shows the result for one specific simulation. The dotted lines show confidence bands based on the simulation results. In addition to the graphical output we calculate the root mean square error (RMSE) and the mean absolute error (MAE) to see how the choice of  $(r, p)$  influences the estimation.



We draw the following conclusions. As already pointed out by Davis and Yau [6] and Huser and Davison [11], there might be a loss in efficiency if too many pairs are included in the estimation. This can be explained by the fact that pairs get more and more independent as the space-time lag increases. Adding more and more pairs to the pairwise log-likelihood function can introduce some noise which decreases the efficiency. This is evident in Figure 6.2 for the temporal parameter  $\alpha_2$ , where the estimates vary more around the mean as more pairs are included in the estimation.

An interesting observation for our model is that using a maximal spatial lag of  $\mathbf{0}$  or a maximal temporal lag 0, respectively, leads to very good results. For the spatial parameters, the space-time lags which lead to the lowest RMSE and MAE are  $(2, 0)$  for  $\theta_1$  and  $(2, 0)$  (RMSE) or  $(3, 0)$  (MAE) for  $\alpha_2$  (see Table 2), i.e. we use all pairs within a spatial distance of 2 or 3 at the same time point. Basically, this suggests that we could also estimate the spatial parameters based on each individual random field for fixed time points and then take the mean over all estimates in time. The same holds for the time parameters  $\theta_2$  and  $\alpha_2$ , where the best results in the sense of the lowest RMSE and MAE are obtained for the space-time lags  $(0, 3)$ , i.e. if we use all pairwise densities corresponding to the space-time pairs  $(\mathbf{s}, t_1)$  and  $(\mathbf{s}, t_2)$ , where  $|t_2 - t_1| \leq 3$  (see Table 3). The reason for this observation is that the parameters of the underlying space-time correlation function get “separated” in the extremal setting in the sense that for example a spatial lag equal to zero does not affect the temporal parameters  $\theta_1$  and  $\alpha_1$  and vice versa.

$\theta_1$	(1,0)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(2,0)	(2,1)	(2,2)
RMSE	0.0123	0.0118	0.0121	0.0122	0.0123	0.0124	0.0103	0.0104	0.0105
MAE	0.0105	0.0090	0.0092	0.0093	0.0094	0.0095	0.0080	0.0081	0.0081
	(2,3)	(2,4)	(2,5)	(3,0)	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)
RMSE	0.0104	0.0104	0.0104	0.0106	0.0107	0.0108	0.0108	0.0107	0.0108
MAE	0.0081	0.0081	0.0081	0.0082	0.0083	0.0083	0.0084	0.0083	0.0084
$\alpha_1$	(2,0)	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(3,0)	(3,1)	(3,2)
RMSE	0.1338	0.1398	0.1530	0.1492	0.1543	0.1569	0.1351	0.1409	0.1579
MAE	0.1078	0.1124	0.1154	0.1137	0.1233	0.1252	0.1050	0.1106	0.1127
	(3,3)	(3,4)	(3,5)	(4,0)	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)
RMSE	0.1596	0.1639	0.1649	0.1423	0.1483	0.1614	0.1673	0.1735	0.1751
MAE	0.1228	0.1291	0.1297	0.1120	0.1176	0.1114	0.1276	0.1372	0.1385

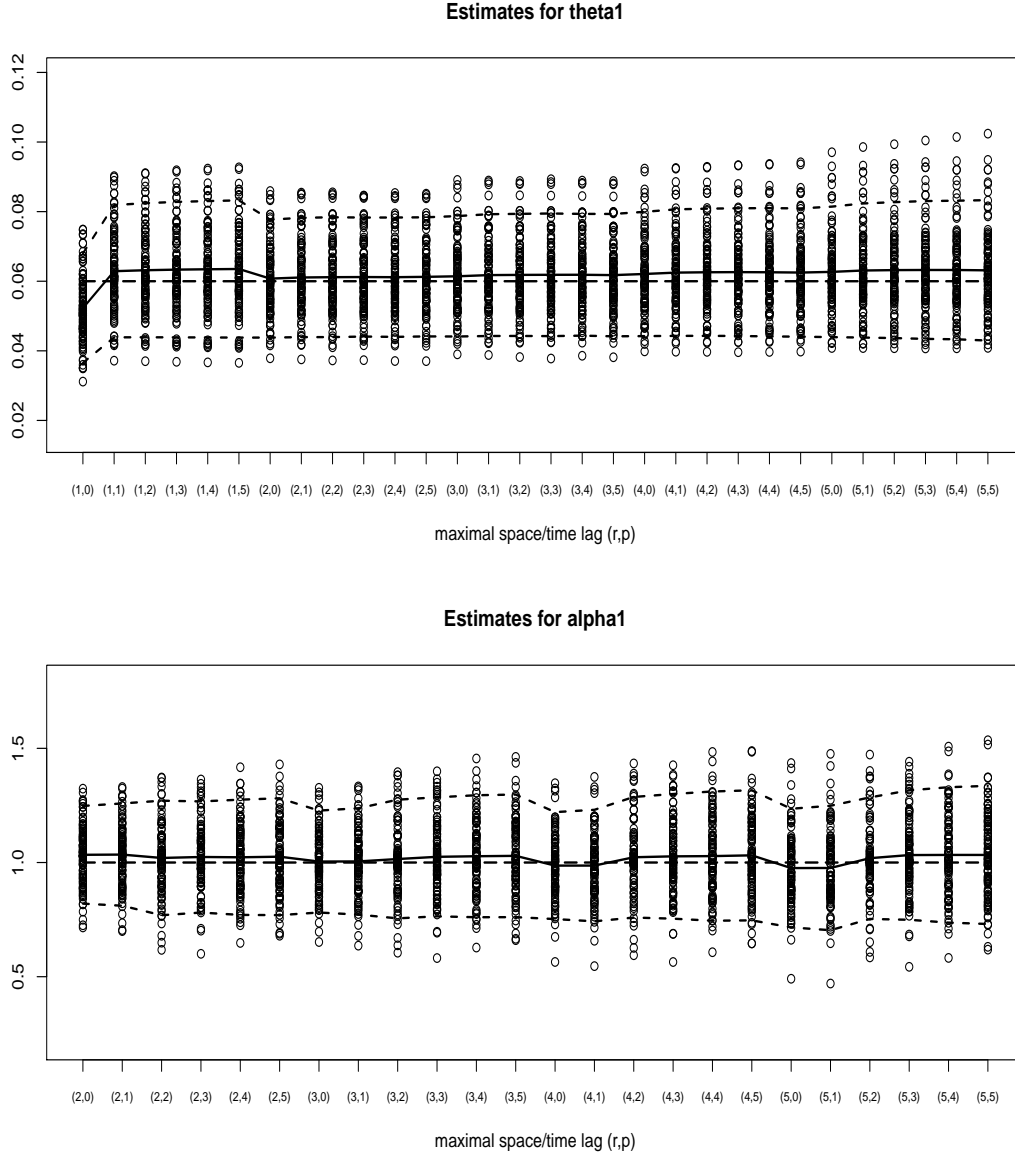
**Table 2:** RMSE and MAE based on 100 simulations for the spatial estimates  $\theta_1$  and  $\alpha_1$  for different combinations of maximal space-time lags  $(r, p)$ .

## Acknowledgments

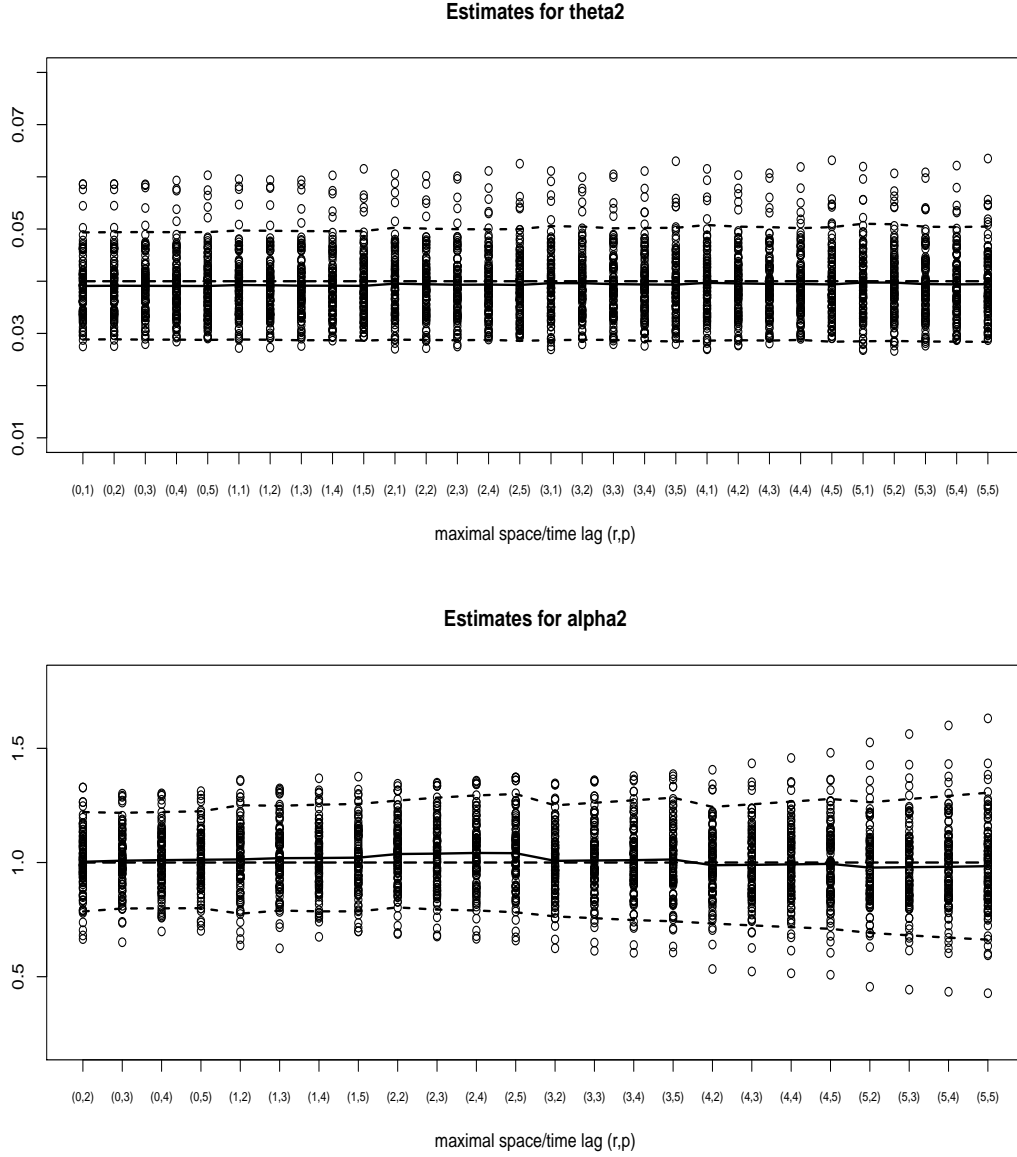
All authors gratefully acknowledge the support by the TUM Institute for Advanced Study (TUM-IAS). The third author additionally likes to thank the International Graduate School of Science and Engineering (IGSSE) of the Technische Universität München for their support.

## References

- [1] J.E. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, 34(2):192–236, 1974.
- [2] E. Bolthausen. On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, 10(4):1047–1050, 1982.



**Figure 6.1:** Estimates for  $\theta_1$  and  $\alpha_1$  (spatial parameters) as a function of maximal space-time lags  $(r, p)$ . Each dot represents the estimate for one of the 100 simulations. The solid line is the mean over all the estimates for each fixed combination of  $r$  and  $p$  and the dotted lines are simulation-based 95% pointwise confidence bands. The middle long dashed line represents the true value.



**Figure 6.2:** Estimates for  $\theta_2$  and  $\alpha_2$  (temporal parameters) as a function of maximal space-time lags  $(r, p)$ . Each red dot represents the estimate for one of the 100 simulations. The solid line is the mean over all the estimates for each fixed combination of  $r$  and  $p$  and the dotted lines are simulation-based 95% pointwise confidence bands. The middle long dashed represents the true value.

$\hat{\theta}_2$	(0,1)	(0,2)	(0,3)	(0,4)	(0,5)	(1,1)	(1,2)	(1,3)	(1,4)
RMSE	0.0182	0.0182	0.0182	0.0182	0.0182	0.0184	0.0183	0.0183	0.0183
MAE	0.0171	0.0171	0.0171	0.0171	0.0171	0.0173	0.0172	0.0171	0.0171
	(1,5)	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(3,1)	(3,2)	(3,3)
RMSE	0.0183	0.0187	0.0186	0.0185	0.0185	0.0185	0.0188	0.0188	0.0186
MAE	0.0171	0.0175	0.0174	0.0173	0.0174	0.0173	0.0176	0.0176	0.0174
$\hat{\alpha}_2$	(0,2)	(0,3)	(0,4)	(0,5)	(1,2)	(1,3)	(1,4)	(1,5)	(2,2)
RMSE	0.1317	0.1269	0.1280	0.1289	0.1442	0.1401	0.1426	0.1438	0.1463
MAE	0.1008	0.0989	0.1015	0.1035	0.1086	0.1079	0.1139	0.1147	0.1179
	(2,3)	(2,4)	(2,5)	(3,2)	(3,3)	(3,4)	(3,5)	(4,2)	(4,3)
RMSE	0.1532	0.1580	0.1619	0.1473	0.1531	0.1589	0.1642	0.1549	0.1607
MAE	0.1242	0.1275	0.1294	0.1169	0.1223	0.1273	0.1317	0.1233	0.1284

**Table 3:** RMSE and MAE based on 100 simulations for the spatial estimates  $\theta_2$  and  $\alpha_2$  for different combinations of maximal space-time lags  $(r, p)$ .

- [3] B.M. Brown and S.I. Resnick. Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739, 1977.
- [4] D.R. Cox and N. Reid. Miscellanea: A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729 – 737, 2004.
- [5] R.A. Davis, C. Klüppelberg, and C. Steinkohl. Max-stable processes for extremes of processes observed in space and time. *Preprint*, 2011. arXiv:1107.4464v1 [stat.ME].
- [6] R.A. Davis and T. Mikosch. Extreme value theory for space-time processes with heavy-tailed distributions. *Stochastic Processes and their Applications*, 118:560–584, 2008.
- [7] R.A. Davis and C.Y. Yau. Comments on pairwise likelihood in time series models. *Statistica Sinica*, 21, 2011.
- [8] L. de Haan. A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194 – 1204, 1984.
- [9] C. Dombry and F. Eyi-Minko. Strong mixing properties of max-infinitely divisible random fields. 2012. arXiv:1201.4645v1.
- [10] T. Gneiting. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 95:590–600, 2002.
- [11] R. Huser and A. Davison. Space-time modelling for extremes. *Preprint*, 2012. arXiv:1201.3245v1 [stat.ME].
- [12] J. Hüsler and R.-D. Reiss. Maxima of normal random vectors: between independence and complete dependence. *Statistics and Probability Letters*, 7:283–286, 1989.
- [13] Z. Kabluchko. Extremes of space-time Gaussian processes. *Stochastic Processes and their Applications*, 119:3962 – 3980, 2009.
- [14] Z. Kabluchko, M. Schlather, and L. de Haan. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37(5):2042 – 2065, 2009.

- [15] U. Krengel. *Ergodic Theorems*. de Gruyter, Berlin, 1985.
- [16] B.G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80:221–239, 1988.
- [17] D. J. Nott and T. Rydén. Pairwise likelihood methods for inference in image models. *Biometrika*, 86(3):661–676, 1999.
- [18] S.A. Padoan, M. Ribatet, and S.A. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association (Theory and Methods)*, 105(489):263–277, 2009.
- [19] M. Schlather. Models for stationary max-stable random fields. *Extremes*, 5(1):33–44, 2002.
- [20] S.A. Stoev. On the ergodicity and mixing of max-stable processes. *Stochastic Processes and their Applications*, 18(9):1679–1705, 2008.
- [21] S.A. Stoev and M.S. Taqqu. Extremal stochastic integrals: a parallel between max-stable processes and  $\alpha$ -stable processes. *Extremes*, 8:237–266, 2005.
- [22] D. Straumann. *Estimation in Conditionally Heteroscedastic Time Series Models*. Lecture Notes in Statistics, Springer, Berlin, 2004.
- [23] C. Varin. On composite marginal likelihoods. *Advances in Statistical Analysis*, 92:1–28, 2007.
- [24] C. Varin and P. Vidoni. Pairwise likelihood inference and model selection. *Biometrika*, 92(3):519–528, 2005.
- [25] A. Wald. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 29:595 – 601, 1946.
- [26] Y. Wang, P. Roy, and S.A. Stoev. Ergodic properties of sum- and max-stable stationary random fields via null and positive group actions. *Technical Report*, 2009.
- [27] Y. Wang and S.A. Stoev. On the structure and representations of max-stable processes. *Advances in Applied Probability*, 42(3):855–877, 2010.

## A Proof of Lemma 5.1

In the following, we use the same abbreviations as in Lemma 3.1. The gradient of the bivariate log-density with respect to the parameter vector  $\psi$  is given by

$$\nabla_{\psi} \log f(x_1, x_2) = \frac{\partial \log f(x_1, x_2)}{\partial \delta} \nabla_{\psi} \delta.$$

Assume in the following that all parameters  $\theta_1, \alpha_1, \theta_2$  and  $\alpha_2$  are identifiable. Since all partial derivatives

$$\frac{\partial \delta}{\partial \theta_1} = \|\mathbf{h}\|^{\alpha_1}, \quad \frac{\partial \delta}{\partial \theta_2} = |u|^{\alpha_2}, \quad \frac{\partial \delta}{\partial \alpha_1} \theta_1 \alpha_1 \|\mathbf{h}\|^{\alpha_1-1}, \quad \frac{\partial \delta}{\partial \alpha_2} = \theta_2 \alpha_2 |u|^{\alpha_2-1},$$

as well as all second order partial derivatives can be bounded from below and above for  $0 < \min \{\|\mathbf{h}\|, |u|\}, \max \{\|\mathbf{h}\|, |u|\} < \infty$  using assumption (21) and, independently of the parameters  $\theta_1, \theta_2, \alpha_1$  and  $\alpha_2$ , it suffices to show that

$$\mathbb{E}_{\psi^*} \left[ \left| \frac{\partial \log f_{\psi}(\eta(\mathbf{s}_1, t_1), \eta(\mathbf{s}_2, t_2))}{\partial \delta} \right|^3 \right] < \infty$$

and

$$\mathbb{E}_{\psi^*} \left[ \sup_{\psi \in \Psi} \left| \frac{\partial^2 \log f_{\psi}(\eta(\mathbf{s}_1, t_1), \eta(\mathbf{s}_2, t_2))}{\partial \delta} \right| \right] < \infty.$$

Since  $\delta$  can be bounded away from zero using assumption (21), we can treat  $\delta$  as a constant. For simplification we drop the argument in the following equalities. Define

$$A_1 = \frac{\partial V}{\partial x_1}, \quad A_2 = \frac{\partial V}{\partial x_2}, \quad \text{and} \quad A_3 = \frac{\partial^2 V}{\partial x_1 \partial x_2}.$$

The partial derivative of the bivariate log-density with respect to  $\delta$  has the following form

$$\frac{\partial \log f_{\psi}}{\partial \delta} = -\frac{\partial V}{\partial \delta} + (A_1 A_2 - A_3)^{-1} \left( \frac{\partial A_1}{\partial \delta} A_2 + A_1 \frac{\partial A_2}{\partial \delta} - \frac{\partial A_3}{\partial \delta} \right).$$

We identify stepwise the “critical” terms, where “critical” means higher order terms of functions of  $x_1$  and  $x_2$ . To give an idea on how to handle the components in the derivatives, we describe one such step. Note that  $(A_1 A_2 - A_3)^{-1}$  can be written as

$$(A_1 A_2 - A_3)^{-1} = \frac{x_1 x_2}{g_1 \left( \frac{1}{x_1}, \frac{1}{x_2}, \frac{1}{x_1 x_2}, \frac{1}{x_1^2}, \frac{1}{x_2^2} \right)},$$

where  $g_1$  describes the sum of the components together with additional multiplicative factors. By using

$$\frac{\partial \Phi(q_{\psi}^{(1)})}{\partial \delta} = \frac{q_{\psi}^{(1)}}{2\delta} \varphi(q_{\psi}^{(1)}) \quad \text{and} \quad \frac{\partial \varphi(q_{\psi}^{(1)})}{\partial \delta} = -\frac{(q_{\psi}^{(1)})^2}{2\delta} \varphi(q_{\psi}^{(1)}),$$

where  $q_{\psi}^{(1)} = \log(x_2/x_1)/(2\sqrt{\delta}) + \sqrt{\delta}$ , we have

$$\frac{\partial A_1}{\partial \delta} A_2 = g_2 \left( \frac{1}{x_1^2 x_2^2}, \frac{q_{\psi}^{(1)}}{x_1^2 x_2^2}, \frac{(q_{\psi}^{(1)})^2}{x_1^2 x_2^2}, \frac{1}{x_1^3 x_2}, \frac{q_{\psi}^{(1)}}{x_1^3 x_2}, \frac{(q_{\psi}^{(1)})^2}{x_1^3 x_2}, \frac{1}{x_1 x_2^3}, \frac{(q_{\psi}^{(1)})^2}{x_1 x_2^3} \right),$$

where  $g_2$  is a linear function of the components. By combining the two representations above, we obtain that all terms in

$(A_1 A_2 - A_3)^{-1} (\partial A_1 / \partial \delta) A_2$  are of the form

$$\frac{|\log x_1|^{k_1} |\log x_2|^{k_2}}{x_1^{k_3} x_2^{k_4}}, \quad k_1, k_2, k_3, k_4 \geq 0. \quad (28)$$

The second derivative of the bivariate log-density with respect to  $\delta$  is given by

$$\begin{aligned} \frac{\partial^2 \log f_{\psi}}{(\partial \delta)^2} = & -\frac{\partial^2 V}{(\partial \delta)^2} - (A_1 A_2 - A_3)^{-2} \left( \frac{\partial A_1}{\partial \delta} A_2 + A_1 \frac{\partial A_2}{\partial \delta} - \frac{\partial A_3}{\partial \delta} \right)^2 \\ & + (A_1 A_2 - A_3)^{-1} \left( \frac{\partial^2 A_1}{(\partial \delta)^2} A_2 + 2 \frac{\partial A_1}{\partial \delta} \frac{\partial A_2}{\partial \delta} + A_1 \frac{\partial^2 A_2}{(\partial \delta)^2} - \frac{\partial^2 A_3}{(\partial \delta)^2} \right) \end{aligned}$$

Stepwise calculation of the single components shows that all terms are also of form (28). This implies that for both statements it suffices to show that for all  $k_1, k_2, k_3, k_4 \geq 0$

$$\mathbb{E} \left[ \frac{(\log \eta(\mathbf{s}, t))^{k_1} (\log \eta(\mathbf{s}, t))^{k_2}}{|\eta(\mathbf{s}, t)|^{k_3} |\eta(\mathbf{s}, t)|^{k_4}} \right] < \infty.$$

Since  $\eta(\mathbf{s}, t)$  is standard Fréchet  $\log(\eta(\mathbf{s}, t))$  is standard Gumbel and  $1/\eta(\mathbf{s}, t)$  is standard exponential. Using Hölder's inequality, we obtain

$$\begin{aligned} & \mathbb{E} \left[ \frac{|\log(\eta(\mathbf{s}, t))|^{k_1} |\log(\eta(\mathbf{s}, t))|^{k_2}}{|\eta(\mathbf{s}, t)|^{k_3} |\eta(\mathbf{s}, t)|^{k_4}} \right] \\ & < \left( \mathbb{E} \left[ |\log(\eta(\mathbf{s}, t))|^{4k_1} \right] \mathbb{E} \left[ |\log(\eta(\mathbf{s}, t))|^{4k_2} \right] \right)^{1/2} \left( \mathbb{E} \left[ \left| \frac{1}{\eta(\mathbf{s}, t)} \right|^{4k_3} \right] \mathbb{E} \left[ \left| \frac{1}{\eta(\mathbf{s}, t)} \right|^{4k_4} \right] \right)^{1/2} < \infty, \end{aligned}$$

since all moments of the exponential and the Gumbel distributions are finite.